

# Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children

Joan A. Sereno, Shari R. Baum, G. Cameron Marean, and Philip Lieberman  
*Brown University, Box 1978, Linguistics Department, Providence, Rhode Island 02912*

(Received 4 April 1986; accepted for publication 15 September 1986)

The present study investigated anticipatory labial coarticulation in the speech of adults and children. CV syllables, composed of [s], [t], and [d] before [i] and [u], were produced by four adult speakers and eight child speakers aged 3–7 years. Each stimulus was computer edited to include only the aperiodic portion of fricative-vowel and stop-vowel syllables. LPC spectra were then computed for each excised segment. Analyses of the effect of the following vowel on the spectral peak associated with the second formant frequency and on the characteristic spectral prominence for each consonant were performed. Perceptual data were obtained by presenting the aperiodic consonantal segments to subjects who were instructed to identify the following vowel as [i] or [u]. Both the acoustic and the perceptual data show strong coarticulatory effects for the adults and comparable, although less consistent, coarticulation in the speech stimuli of the children. The results are discussed in terms of the articulatory and perceptual aspects of coarticulation in language learning.

PACS numbers: 43.71.An, 43.70.Bk, 43.71.Es, 43.70.Fq

## INTRODUCTION

In the production of speech, the acoustic properties of certain sounds are changed under the influence of adjacent sounds. These coarticulation effects are categorized as anticipatory or “forward” coarticulation and carryover or “backward” coarticulation. Whereas carryover coarticulation is often attributed to the inertial properties of the articulators, anticipatory coarticulation appears to result from some sort of preplanning stage.

One of the most extensively studied anticipatory coarticulation effects has been the rounding of consonants preceding a rounded vowel. The present study examines labial anticipatory coarticulation in the speech of adults and children by means of an acoustic analysis of stimuli produced by adults and children and the perceptual assessment of these utterances.

Many articulatory studies using a variety of techniques have demonstrated the existence of labial coarticulation in adults. Daniloff and Moll (1968), as well as Lubker (1981), citing an earlier study by Lubker *et al.* (1975), found that anticipatory lip rounding, measured by the amount of lip protrusion using cinefluorographic film and the amount of electromyographic (EMG) activity of the orbicularis oris muscle, begins in the first consonant of a series of consonants that precede the rounded vowel [u]. This articulatory effect has also been corroborated in a study of French utterances, in which lip protrusion occurred in the first of a series of four to six consonants preceding a rounded vowel (Benguerel and Cowan, 1974).

In contrast, other data (Gay, 1978, 1979; Bell-Berti and Harris, 1979) seem to suggest that anticipatory coarticulation is not completely regulated by the number of preceding unrounded consonants but rather occurs at a fairly constant time interval prior to the onset of the rounded vowel. When a greater number of consonants increased total consonant du-

ration, there was no comparable increase in the onset time of muscle activity relative to the start of the vowel.

In an attempt to specifically examine these two opposing viewpoints of coarticulation (time variable versus time locked), Lubker and Gay (1982) systematically manipulated possible confounding experimental, biological, and linguistic variables that may have produced the contradictory data. Language-specific differences, such as the need to emphasize particular rounding distinctions (in Swedish, for example, but not in English), occurred. Lubker and Gay suggest that significant language-specific differences have an effect on the underlying processes involved in anticipatory labial coarticulation. The data also show that the control of anticipatory labial coarticulation differs not only across various languages, but also from speaker to speaker. These results argue against the possibility that anticipatory coarticulation derives from uniform, innate, genetically transmitted motor control patterns. Although individual differences might conceivably be genetically transmitted, it is most implausible that language-specific differences, for example between Swedish and English, are genetically transmitted.

Although anticipatory coarticulation of lip rounding is articulatorily observable at a distance from the rounded vowel, studies of English (Lehiste and Shockey, 1972) and French (Benguerel and Adelman, 1975) showed no comparable *perceptual* effect of coarticulation in nonadjacent segments. Even when the first segment of a consonant cluster contained information about the articulation of the upcoming rounded vowel, listeners required at least the consonant immediately preceding the vowel to correctly identify that vowel. In investigations of the perceptual effects of coarticulation in stop-vowel monosyllables, Winitz *et al.* (1972), LaRiviere *et al.* (1975a), and Ohde and Sharf (1977) found that subjects, given only the brief aperiodic section of the preceding consonantal segment, were capable of correctly

identifying the “missing” vowel. These results were corroborated by Blumstein and Stevens (1980) using synthetic CV syllables. From these studies, then, it would seem that information regarding the succeeding vowel is not only acoustically present in the preceding consonants but is also a perceptually salient cue for the identification of that following vowel.

In comparable studies of fricative-vowel monosyllables, LaRiviere *et al.* (1975b) reported that when only the frication noise spectra of voiceless fricatives are given, cues for vowel recognition are present in 8 of the 12 stimuli (four fricatives [f, θ, s, ʃ] in three different vowel contexts [i, a, u]). More recently, Yeni-Komshian and Soli (1981) confirmed that anticipatory coarticulation in frication segments provides perceptual cues for excised vowels. Although two vowels ([i] and [u]) were more easily identified than a third ([a]), all vowel identification scores exceeded chance level (33%) in each fricative context.

These articulatory and perceptual data on coarticulation are also consistent with many studies examining the acoustic effect of a vowel environment on a preceding consonant (e.g., Liberman *et al.*, 1967). Ohman (1966) demonstrated that the consonants in VCV utterances depend primarily on the final vowel context. The articulatory adjustment of the consonant in anticipation of the following vowel resulted in systematic acoustic differences in utterances from Swedish, American, and Russian speakers. Fant (1973) presents further cross-linguistic data from English and Swedish showing a range of stop formant frequencies ( $F_2$  and  $F_3$ ) dependent upon the following vowel.

In addition, a spectral analysis of fricative segments also clearly showed anticipatory coarticulation effects (Soli, 1981). The second formant frequencies of all the sibilant fricatives were 100–300 Hz higher before the front vowel [i] compared to the back vowels [a, u]. These acoustic manifestations of anticipatory coarticulation were present in the frication segments about 30–60 ms before the vowel’s first pitch period. It seems, then, that brief initial consonantal segments excised from both stop-vowel and fricative-vowel syllables display systematic acoustic differences in their spectral properties and provide sufficient cues for the appropriate perceptual identification of the following vowel.

Previous articulatory, perceptual, and acoustic studies, therefore, have found anticipatory labial coarticulatory cues in adult speech. However, children’s speech in general shows significantly more variability. Recent studies (Kent, 1983; Sharkey and Folkins, 1985; Watkin and Fromm, 1984) have found that young children display significant variability in a variety of motor processes for speech, suggesting a late development for coarticulatory processes. However, in two recent studies, Turnbaugh *et al.* (1985), using stop-vowel-stop syllables, and Nittrouer (1985), using fricative-vowel words, found that children seem to coarticulate more than adults. In both studies, second formant values were lowered more for consonants preceding [u] in children’s speech compared to adult speech. Nittrouer (1985) claimed that these results support an acquisition process in which children’s initial analysis of phonological sequences may be in terms of syllabic units which contain coarticulatory infor-

mation rather than segments which do not (Menyuk and Menn, 1979).

The objective of the present study is to examine further possible labial coarticulation differences between adults and children. Since many previous studies have shown that there is greater articulatory variability in the speech of children (e.g., Sharkey and Folkins, 1985), and that coarticulatory lip rounding is not uniform with respect to different speakers or different languages, it is reasonable to suppose that anticipatory coarticulation may be acquired. If this is the case, then strong anticipatory labial coarticulatory effects would be both acoustically observable and perceptually present in adult speech and either variable or nonexistent in the speech of children.

In the present study, one fricative consonant [s] and two stop consonants [t, d] in two vowel environments [i, u] were analyzed acoustically from the speech of four adults and eight children. Further experiments were then performed to determine whether any coarticulatory interactions found were perceptually salient.

## I. EXPERIMENT 1

Experiment 1 consisted of an acoustic analysis of CV syllables to determine the effects of anticipatory labial coarticulation on the initial consonantal segment.

### A. Methods

#### 1. Stimuli

Tokens of six different CV utterances [si, su, ti, tu, di, du] were produced in isolation by four adult and eight child speakers of American English. The children ranged in age from 3–7 years old. Each utterance was read from a 3 × 5 card by the adult subjects and repeated by the children. All speakers pronounced the syllables in the order [si, su, ti, tu, di, du] and then the entire sequence was repeated. The syllables were recorded on magnetic tape in a sound-treated room with a Nagra 4.2 tape recorder and a Shure SM81 microphone. Stimuli were then transferred to a PDP-11/34 computer using a 20-kHz sampling rate and a 9.0-kHz low-pass filter setting.

Each of the four adult speakers produced five repetitions of every token [si, su, ti, tu, di, du]. Each of the eight children produced between three and six repetitions of each token (due to occasional uncooperativeness on the part of some children). Only those child tokens judged by the experimenters to be highly intelligible were included in the analyses; the actual numbers of tokens used are listed in Table I with the acoustic results. For the present analyses, then, there were 120 adult stimuli and 182 child stimuli.

The waveform of each stimulus was visually displayed for editing. For the fricative stimuli, the frication noise was defined as the segment from the onset of the utterance up to the onset of periodicity. For the stop consonant stimuli, the aperiodic portion of the signal was defined also as the segment immediately prior to the onset of periodicity. This segment consisted of a burst and aspiration noise. Thus neither the frication noise of the fricative nor the aperiodic portion of the stop consonants contained any vowel transitions.

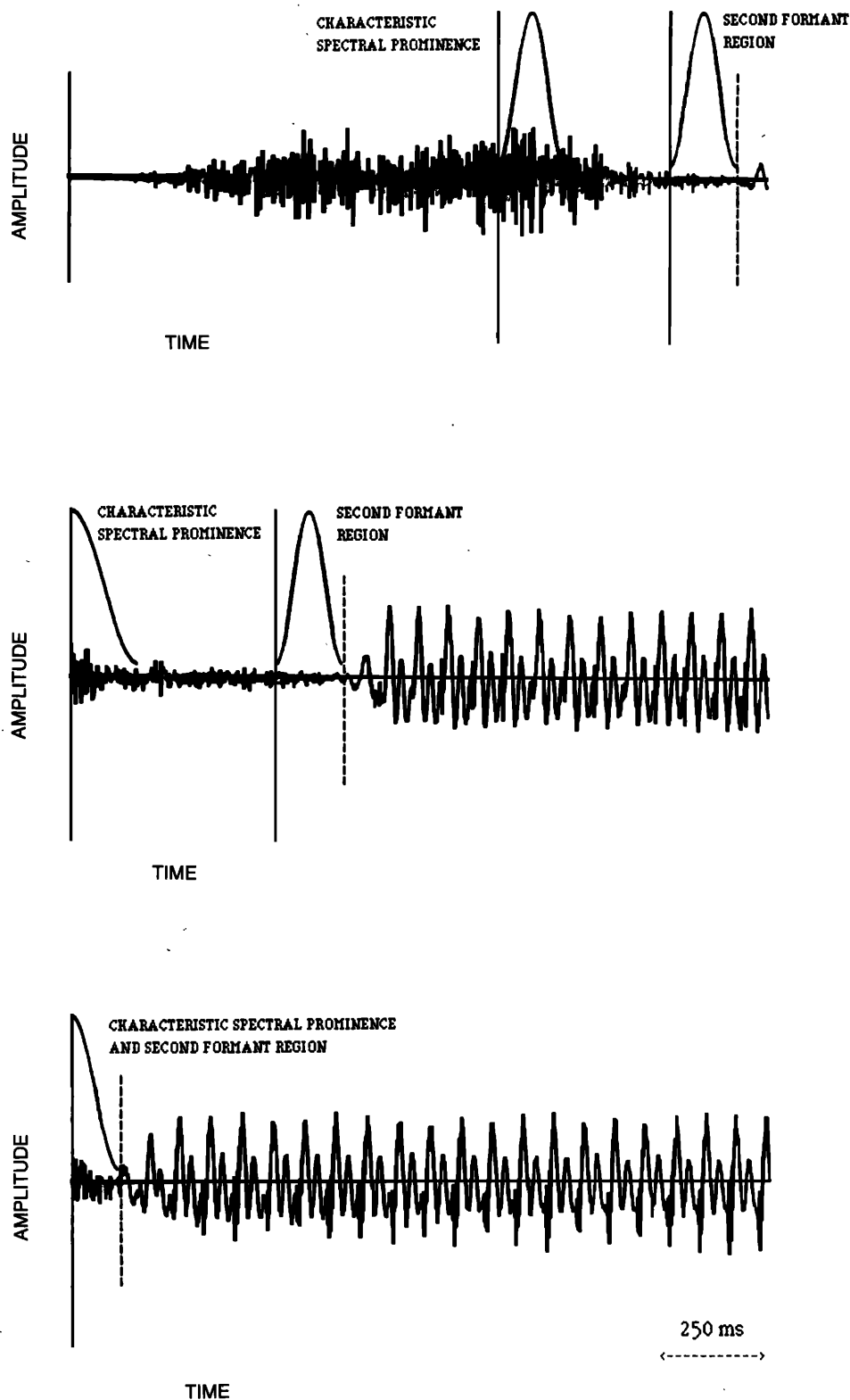


FIG. 1. Waveform displays of [s] (top panel), [t] (middle panel), and [d] (bottom panel) stimuli are shown with superimposed Hamming windows used for sampling the spectrum at both the characteristic spectral prominence and the second formant region.

For the adult stimuli, the average fricative duration for the fricative-vowel stimuli was 240 ms, whereas the average aperiodic consonant duration was 99 ms for unvoiced stops and 18 ms for voiced stops. For the child stimuli, the average fricative duration for the fricative-vowel stimuli was only 195 ms, whereas the average aperiodic consonant duration was 104 ms for unvoiced stops and 29 ms for voiced stops.

## 2. Procedure

Spectra of the aperiodic segments were derived by means of computer-implemented 24-term linear predictive coding (LPC) and Fourier analysis (see Fig. 1). For the fricative segments [si] and [su], a 20-ms full Hamming window was placed at two different points: the final 20 ms of the fricative noise and 70 ms prior to the onset of the vowel

TABLE I. Mean spectral peak values measured at the characteristic spectral prominence and at the second formant region for both adults and children. Asterisks (\*) represent measurements that were in the unexpected direction. Number of tokens analyzed are in parentheses.

Adults						
Characteristic spectral prominence						
CV syllables						
Speaker	[si]	[su]	[ti]	[tu]	[di]	[du]
A1	4090	3460	3842	3598	3796	3528
A2	4712	4661	4134	4240*	3599	3969*
A3	5160	4798	4795	3371	3745	3625
A4	4664	4183	3972	3613	3643	3362
Mean	4657	4276	4186	3706	3696	3621
Second formant region						
CV syllables						
Speaker	[si]	[su]	[ti]	[tu]	[di]	[du]
A1	1747	1584	2106	1546	2512	1825
A2	2302	1996	2634	2009	2292	2473*
A3	1905	1757	2087	1796	2045	1753
A4	1834	1647	2000	1687	1909	1739
Mean	1947	1746	2207	1760	2190	1948
Children						
Characteristic spectral prominence						
CV syllables						
Speaker (age in months)	[si]	[su]	[ti]	[tu]	[di]	[du]
C1(36)	5738(1)	5426(2)	4888(3)	4811(3)	4849(3)	5023(2)*
C2(47)	6433(2)	6272(2)	5538(2)	5069(2)	5360(2)	5159(2)
C3(61)	6484(3)	6264(6)	5986(6)	5530(5)	5521(6)	5686(5)*
C4(66)	5798(2)	5745(2)	5284(3)	5182(2)	4496(3)	4600(3)*
C5(66)	5723(4)	5293(4)	5299(4)	5222(2)	4857(1)	4586(3)
C6(68)	6476(4)	5717(1)	5951(5)	5886(4)	5884(4)	5819(5)
C7(77)	5002(4)	4580(4)	4723(5)	5094(3)*	4190(4)	3917(5)
C8(85)	6089(4)	5750(4)	5911(4)	5668(3)	6097(4)	6008(3)
Mean	5968	5631	5448	5308	5157	5100
Second formant region						
CV syllables						
Speaker (age in months)	[si]	[su]	[ti]	[tu]	[di]	[du]
C1(36)	3470(1)	3067(2)	3401(3)	3632(3)*	3264(3)	3182(2)
C2(47)	2648(2)	2291(2)	2910(2)	2404(2)	3022(2)	2939(2)
C3(61)	2469(3)	2358(6)	2939(6)	2624(5)	1540(5)	1573(5)*
C4(66)	2314(2)	2335(3)*	3227(2)	2556(3)	2912(3)	2416(3)
C6(66)	2096(3)	2020(2)	2714(4)	2295(2)	2723(1)	2591(1)
C6(68)	2645(4)	2627(1)	2864(5)	2558(4)	2719(4)	2654(4)
C7(77)	2646(5)	2066(3)	3081(5)	2166(3)	2880(4)	2471(5)
C8(85)	2702(3)	2041(4)	2903(4)	2505(4)	2808(3)	2703(4)
Mean	2624	2351	3005	2593	2734	2566

transition (i.e., a point of relatively high amplitude). In the spectral analysis of the final 20 ms of the frication noise, the major spectral peak was measured between 1.5 and 3.0 kHz for the adult stimuli and between 1.5 and 3.5 kHz for the child stimuli in anticipation of the second formant of the following vowel. This spectral area will be referred to as the second formant region of the consonant. In the spectral anal-

ysis of the frication noise 70 ms prior to the onset of the vowel transition, the spectral peak was measured between 3.0 and 5.0 kHz for the adult stimuli and between 4.5 and 6.5 kHz for the child stimuli, consistent with the major distinctive peak in the [s] spectrum noted by Heinz and Stevens (1961) and Hughes and Halle (1956). This region will be referred to as the characteristic prominence for [s]. It should be noted that the spectra for children are, not unexpectedly, shifted substantially upwards from those for adults because of the shorter length of a young child's supralaryngeal vocal tract (Lieberman, 1980).

For the unvoiced stop consonant segments [ti] and [tu], a 20-ms half Hamming window was placed over the initial 20 ms of the utterance and then a full Hamming window was displayed over the final 20 ms of the aperiodic portion of the stimulus. Major spectral peaks were measured in the same frequency ranges as for [s] for the adult and child, respectively.

Since the voice onset time (VOT) for voiced stop consonants is considerably shorter than for their voiceless counterparts, two distinct areas of the aperiodic portion of the voiced stop consonant segments [di] and [du] could not be analyzed using a 20-ms window. Instead, a variable half Hamming window was placed over the entire aperiodic portion of the stimulus. Again, the same two frequency regions were examined.

## B. Results

Mean spectral peak values at the two frequency regions for the three consonants in the two vowel environments were calculated for each of the adults' and children's stimuli. These values are displayed in Table I. In all cases, overall mean values for consonants in the context of [i] are higher than those for consonants in the context of [u].

We will first consider values measured at the characteristic prominence frequency region. For the [si] stimuli at the characteristic prominence, the overall mean for the adult subjects was 4657 Hz, whereas that for [su] stimuli was 4276 Hz, a difference of 381 Hz. For the adult [ti] and [tu] stimuli, the difference was again great—4186 vs 3706 Hz—a difference of 480 Hz. The overall difference between the [di] and [du] stimuli was much smaller at 75 Hz (3696 vs 3621 Hz). The children's data paralleled those of the adults with the mean characteristic prominence for [si] stimuli falling at 5968 Hz and that for [su], 5631 Hz—a difference of 337 Hz; [ti] and [tu] measurements differed by 140 Hz (5448 vs 5308 Hz), whereas the difference between [di] and [du] at the characteristic spectral prominence region was 57 Hz (5157 vs 5100 Hz).

Values measured at the second formant region demonstrated a similar pattern of results, with spectral peak values for consonants preceding [i] higher, on average, than those for consonants preceding [u] for both the adults and the children. For the [si] stimuli in this second formant region, the overall mean for the adult subjects was 1947 Hz, whereas that for [su] was 1746 Hz, a difference of 201 Hz. The mean difference between the [ti] and [tu] stimuli was 447 Hz (2207 vs 1760 Hz); that between [di] and [du] was 242 Hz (2190 vs 1948 Hz). The children's data showed similar dif-

ferences measured in the second formant frequency range. Mean peak value for [si] stimuli was 2624 Hz with that for [su] stimuli at 2351 Hz. For [ti] and [tu], the values measured differed by considerably more, with the mean for [ti] tokens at 3005 Hz and that for [tu] at 2593 Hz (a difference of 412 Hz). The means for the [di] and [du] stimuli produced by the children differed by only 168 Hz, ([di] 2734 Hz versus [du] 2566 Hz).

The mean frequency values were submitted to an analysis of variance (ANOVA) comparing group (adult versus child), frequency range (characteristic spectral prominence versus second formant), consonant ([s] vs [t] vs [d]), and vowel ([i] vs [u]). Results of the ANOVA revealed significant main effects for group [ $F(1,10) = 47.31, p < 0.001$ ], frequency range [ $F(1,10) = 203.22, p < 0.001$ ], consonant [ $F(2,20) = 5.62, p < 0.05$ ], and vowel [ $F(1,10) = 48.95, p < 0.001$ ], as expected. The most interesting effect is that for vowel, since it indicates that, overall, values calculated for consonants preceding [i] differed significantly from those for consonants in the environment of [u], the predicted effect of coarticulation of lip rounding.

No significant interactions were found between group and vowel [ $F(1,10) = 0.92, p < 0.3$ ], nor between group, vowel, and frequency range [ $F(1,10) = 0.55, p < 0.4$ ], suggesting that the adults and children did not differ with respect to the influence of the rounded vowel on their consonant production. That is, both adults and children demonstrate a significant lowering of spectral energy in both frequency ranges examined (or, equivalently, both points in time), for consonants produced in the environment preceding [u] as compared to [i]. This was true for all three consonants examined.

However, if one looks again at Table I, which presents the mean peak values for each subject, there appears to be greater variability in the children's data. That is, not all frequency values pattern in the expected direction in the children's data. This is especially true of speakers C1, C3, C4, and C7, for which inconsistent peak values were calculated. More will be said about this increased variability as it relates to the perceptual findings.

### C. Summary of results

The results of our acoustic analyses indicate that both adults and children aged 3–7 years demonstrate an acoustic effect of coarticulation of lip rounding. For both speaker groups, consonants produced in the environment preceding [u] displayed significantly lower spectral energy peaks than those produced before [i], even at the onset of stop stimuli and 70 ms prior to vowel onset for the fricative stimuli. There was, however, more individual variability in the children's data, as might be expected. In the following experiment, these CV utterances were submitted to a perceptual analysis.

## II. EXPERIMENT 2

Experiment 2 consisted of a perceptual study in which the aperiodic portions corresponding to the consonants of CV syllables were presented to subjects to determine whether the acoustic manifestations of coarticulation are perceptually salient to naive listeners.

## A. Methods

### 1. Subjects

Ten students at Brown University were paid for their participation in the perceptual experiment. All subjects were native speakers of English with no known hearing impairments.

### 2. Materials

The entire aperiodic portion, as described earlier, was excised from each of the consonant–vowel stimuli. A tape recording of these aperiodic stimuli was then made, with the adult stimuli preceding the children's stimuli. Both the adults' and children's stimuli were presented in separate blocks—first the [s] stimuli, then the [t] stimuli, and finally the [d] stimuli. Within each block, every stimulus item appeared twice. There was an ISI of 3 s with a 6-s interval between every ten stimuli. All stimuli were randomized within each block. Therefore, subjects heard a tape recording consisting of 240 adult stimuli followed immediately by 364 child stimuli.

### 3. Procedure

Stimuli were presented binaurally on a Technics 1500 reel-to-reel tape deck with a McIntosh MC2120 amplifier using AKG K141 headphones. No more than three subjects were tested at any one time. At least ten practice items were first presented in order to accustom the subjects to the task. No feedback was provided. Each subject was instructed to identify in a forced-choice paradigm the absent vowel, [i] or [u], given the aperiodic portion of the presented stimuli.

## B. Results

Responses of each subject were scored for percent correct vowel identification, for each of the three consonants produced by the adults and the children. Mean correct identification scores for the adults' stimuli are 68%, 88%, and 67% for [s], [t], and [d], respectively. For the children's tokens, mean scores were lower, at 55% for [s], 83% for [t], and 60% for [d] tokens. Table II provides the mean percent correct scores for each subject for both the adults' and chil-

TABLE II. Percent correct identification of the vowels [i] and [u] for the [s], [t], and [d] stimuli of both adult and child speakers. Asterisks (\*) represent mean scores significantly above chance.

Subject	Adults			Children		
	[s]	[t]	[d]	[s]	[t]	[d]
S1	64	84	51	45	84	52
S2	69	94	74	64	81	59
S3	66	85	69	62	85	62
S4	68	83	73	50	85	63
S5	80	94	63	66	92	69
S6	65	90	70	50	88	63
S7	79	94	65	61	84	66
S8	65	94	68	56	84	59
S9	66	89	73	49	75	47
S10	60	74	59	49	75	56
Mean	68*	88*	67*	55	83*	60

dren's stimuli. It is clear from this table that the [t] tokens consistently yield the best identification scores for both the adults' and children's stimuli. All perception scores for the adults' stimuli were found to be significantly better than chance (binomial distribution,  $p < 0.05$ ), whereas only that for the [t] stimuli produced by the children was significantly better than chance. In addition, a group (adult versus child) by consonant ([s] vs [t] vs [d]) ANOVA was conducted to compare the perception scores. This revealed a significant main effect for group [ $F(1,9) = 32.00, p < 0.001$ ] and for consonant [ $F(2,18) = 152.08, p < 0.001$ ]. A reliable group  $\times$  consonant interaction was also found [ $F(2,18) = 4.09, p < 0.05$ ], suggesting that perception scores for the adults' and children's stimuli differed by consonant. *Posthoc* Newman-Keuls tests demonstrated significant differences ( $p < 0.05$ ) between each of the adults' and children's individual consonants; e.g., scores for the adults' [s] productions were significantly better than those for the children's [s] productions. In addition, perception scores for the adults' [t] stimuli differed significantly from both [s] and [d] scores; a similar pattern was found for the children's tokens.

Errors on the perception test were broken down by speaker for further analysis. For the adults, speaker A2 yielded the highest number of errors in vowel identification for all three consonants, with correct identification scores of 54%, 76%, and 49% for [s], [t], and [d], respectively. These data seem to correspond to the less consistent acoustic cues found for this subject. For the children's data, there was considerably more variability. Syllables produced by speakers C1, C3, C4, and C7 resulted in the highest error rates across all consonants. Even for the [t] stimuli, for which overall correct identification scores were fairly high at 83%, correct identification for C1, C3, C4, and C7 was only 78% with that for C2, C5, C6, and C8 much higher at 89%. These data, too, agree well with the acoustic results reported earlier.

Overall, then, subjects could fairly accurately identify the vowel as [i] or [u] given only the aperiodic portion of the adults' syllables. Identification scores for the children's stimuli were significantly lower than those for the adults' tokens for all three consonants.

### III. DISCUSSION

The present study investigated the acoustic effects and perceptual salience of anticipatory labial coarticulation in the speech of adults and children. As for the adult data, these experiments show that the acoustic effects of coarticulation are congruent with the results of the perceptual test, in which listeners were asked to identify the absent vowel. In cases where the context of the following vowel resulted in a substantial shift in the spectrum of the consonantal stimuli (both at the characteristic spectral prominence of the consonant and at the region approximating the frequency position of the second formant of the following vowel), there was also a significant correct perceptual identification of the absent vowel. Subjects could identify the vowel from adult [s], [t], and [d] tokens with 68%, 88%, and 66% correct identification scores, respectively. The spectra of these stimuli were

shifted on average 312 Hz lower at the characteristic spectral prominence, and 297 Hz at the second formant region in the context of [u] compared to [i]. The presence of a rounded vowel following a consonant resulted in a displacement of the entire spectrum for that consonant. For adult speech, then, the acoustic analysis shows systematic vowel context effects, and the perceptual data show that listeners' perception of these coarticulatory effects is consistent with the acoustic analysis. Listeners apparently can utilize these coarticulatory cues.

Some of the children's data also seem to demonstrate this congruence between the acoustic data and the perceptual results. For the children's [t] tokens, there was on average a 276-Hz spectral shift with a corresponding perceptual identification score of 83% correct. While the acoustic evidence showed a less substantial spectral lowering for the children (276 Hz) compared to the adults (464 Hz), there was also a comparable lowering of the perceptual identification scores (children—83% correct; adults—88% correct). Nevertheless, the analysis of the children's [t] tokens does reveal relatively large acoustic changes due to labial coarticulation and comparable perceptual identification of these cues.

The children's [d] tokens show a similar patterning in that the acoustic data match the perceptual results. However, for the children's [d] stimuli, there does not seem to be either acoustic or perceptual evidence for anticipatory labial coarticulation. There was on average only a 112 Hz lowering of [d] segments in the context of [u] compared to the context of [i] and a corresponding nonsignificant perceptual identification score of only 60% correct. Children show only a small acoustic effect of labial coarticulation in their [d] stimuli, and these minor differences are not perceptually salient.

For the children's [s] tokens, however, the acoustic analysis does not match the perceptual results. While the spectral analysis revealed a strong downward shift for the [s] stimuli in the context of [u] both at the spectral prominence (371 Hz) and at the second formant region (273 Hz), listeners were not able to make use of these cues. The subsequent vowel could only be identified at chance level (55% correct identification). These results seem to suggest that at least for the children's tokens, although a spectral lowering does exist in [s] stimuli preceding [u], the measured acoustic effect is not perceptually salient.

These data are significant in view of the previous studies examining coarticulatory effects in children. For example, Turnbaugh *et al.* (1985) only used acoustic measures, whereas Nittrouer (1985), in addition, investigated children's perception of synthetic stimuli. Her study suggests that children may be sensitive to conflicting vowel context cues and vocalic transition information, but does not directly address whether vocalic information is identifiable in naturally produced consonantal stimuli. Our data suggest that, although a robust acoustic effect may be observable in the children's stimuli, it is not clear that these acoustic cues are always perceptually salient; perhaps these acoustic manifestations are not those which provide listeners with coarticulatory cues.

The present study also illustrates the difference between the adults' and children's data. This distinction becomes most clear in examining the perceptual data. In the adult consonantal stimuli, the vowel was accurately perceived over 74% of the time, with all consonantal contexts ([s], [t], and [d]) having identification scores significantly above chance level. The perceptual results of the children's utterances, in contrast, show a less consistent coarticulatory effect. For the children's stimuli, there was an overall perceptual identification score of just 66% correct, significantly lower than that for the adults' stimuli. However, this perceptual effect is almost entirely the result of coarticulatory cues present in the children's [t] stimuli with a mean perceptual identification score for these stimuli of 83% correct, and scores for the other two consonants ([s] and [d]) at chance level (58%).

These perceptual results suggest that anticipatory labial coarticulation may constitute a generalizable change beginning in unvoiced alveolar stops and spreading to other consonants. Several researchers have found evidence of similar processes in the acquisition of various phonemes (cf. Chapin-Ringo, 1985; Macken and Barton, 1980). For example, Chapin-Ringo (1985) reports a gradual acquisition of /st/ clusters followed by a very rapid acquisition of /sp/ and /sk/ clusters demonstrating generalization to a phonological class of /s/ + stop clusters. The present data are comparable in showing that coarticulatory effects are perceptually salient in the child's [t] stimuli but are either variable or nonexistent in the other consonantal segments.

The results of the present study may be interpreted in terms of recent data on the development of patterned motor activity. The general process of automatization or motor programming underlies most motor behaviors. Automatization yields rapid goal-directed responses to external stimuli in the absence of specific innate mechanisms (Evarts, 1973; Polit and Bizzi, 1978; Miles and Evarts, 1979). One of the most highly complex motor behaviors is the coordination of articulation in human speech. Observations of normal speech muscle activity suggest that these actions, too, can be understood as complex goal-directed maneuvers (Abbs and Gracco, 1984; Gracco and Abbs, 1985; Lieberman, 1984, 1985), involving varying patterns of muscle actions which are completely subordinate to the acoustic goals of the entire speech gesture (Abbs, 1986).

Recent data indicate that the attainment of goal-directed motor activity (involving the lips) takes many years. Watkin and Fromm (1984) showed that the variability of both upper and lower lip displacement in the production of vowels decreased in children from 4 to 10 years of age, with 4- and 7-year-old children showing significantly more variability than the older children or adults. Using additional measures, Sharkey and Folkins (1985) found that, while there was no significant difference in variability across child groups (ages 4, 7, and 10 years old), adult stimuli were significantly less variable than any of the children's utterances. However, measures of lip displacement in the young children (age 4) showed significantly more variability than either the older children (age 7 and 10) or the adults, suggesting different age levels for the acquisition of individual

motor processes for speech. All these results point to a period within the range of 4 to 10 years during which speech motor patterns are intensively refined.

The present acoustic analyses support the conclusion that children's utterances (in this case, those of children between ages 3 and 7 years) exhibit less precise, more variable coarticulatory effects than adult utterances. More important, the perceptual tests in which subjects were required to identify the vowel provide corroborating evidence that the acoustic cues convey the coarticulatory information. These data, furthermore, show that the child does not generalize anticipatory coarticulation across all consonants, a result which is consistent with models of acquisition in which the child initially starts on a word-by-word or phoneme-by-phoneme basis, and only later generalizes across phonetic features or classes of phonemes.

The differences that we found between children's and adults' speech may reflect the development of automatized speech motor control patterns. The data suggest that the realization of the motor programs that underlie anticipatory coarticulation is not innate. Even for lip rounding, there are differences depending on the nature of the segmental elements involved. The results are consistent with a developmental process involving gradual acquisition and fine tuning of speech motor patterns.

## ACKNOWLEDGMENTS

This research was supported by funds from the John D. and Catherine T. MacArthur Foundation Research Network on the Transition from Infancy to Early Childhood and by an NSF Graduate Fellowship to the first author. We would also like to express our appreciation to the Harvard Infant Study Center for assistance in data collection and to Joanne Miller for helpful suggestions on an earlier draft.

- Abbs, J. (1986). "Invariance and variability in speech production: a distinction between linguistic intent and its neuromotor implementation," in *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt (LEA, Boston).
- Abbs, J., and Gracco, V. (1984). "Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech," *J. Neurophysiol.* **51**, 705-723.
- Bell-Berti, F., and Harris, K. (1979). "Anticipatory coarticulation: Some implications from a study of lip rounding," *J. Acoust. Soc. Am.* **65**, 1268-1270.
- Benguerel, A.-P., and Adelman, S. (1975). "Coarticulation of lip rounding and its perception," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. Nooteboom (Springer, New York).
- Benguerel, A.-P., and Cowan, H. A. (1974). "Coarticulation of upper lip protrusion in French," *Phonetica* **30**, 41-55.
- Blumstein, S., and Stevens, K. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.* **67**, 648-662.
- Chapin-Ringo, Carol. (1985). "The nature of change in phonological development: evidence from the acquisition of /s/ + stop and /s/ + nasal clusters," Doctoral dissertation, Brown University, Rhode Island.
- Daniiloff, R., and Moll, K. (1968). "Coarticulation of lip rounding," *J. Speech Hear. Res.* **11**, 706-721.
- Evarts, E. (1973). "Motor cortex reflexes association with learned movement," *Science* **179**, 501-503.
- Fant, G. (1973). "Stops in CV-syllables," in *Speech Sounds and Features*, edited by G. Fant (MIT, Cambridge).

- Gay, T. (1978). "Articulatory units: segments or syllables?," in *Syllables and Segments*, edited by A. Bell and J. Hooper (North-Holland, Amsterdam).
- Gay, T. (1979). "Coarticulation in CV syllables," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, New York).
- Gracco, V., and Abbs, J. (1985). "Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes," *J. Neurophysiol.* **54**, 418–432.
- Heinz, J., and Stevens, K. (1961). "On the properties of voiceless fricative consonants," *J. Acoust. Soc. Am.* **33**, 589–596.
- Hughes, G., and Halle, M. (1956). "Spectral properties of fricative consonants," *J. Acoust. Soc. Am.* **28**, 303–310.
- Kent, R. D. (1983). "The segmental organization of speech," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York).
- LaRiviere, C., Winitz, H., and Herriman, E. (1975a). "Vocalic transitions in the perception of voiceless initial stops," *J. Acoust. Soc. Am.* **57**, 470–475.
- LaRiviere, C., Winitz, H., and Herriman, E. (1975b). "The distribution of perceptual cues in English prevocalic fricatives," *J. Speech Hear. Res.* **18**, 613–622.
- Lehiste, I., and Shockey, L. (1972). "On the perception of coarticulation effects in English VCV syllables," *J. Speech Hear. Res.* **15**, 500–506.
- Liberman, A., Cooper, F. S., Shankweiler, D., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Lieberman, P. (1980). "On the development of vowel production in young children," in *Child Phonology: Perception and Production*, edited by G. Yeni-Komshian and J. Kavanagh (Academic, New York).
- Lieberman, P. (1984). *The Biology and Evolution of Language* (Harvard U. P., Cambridge).
- Lieberman, P. (1985). "On the evolution of human syntactic ability. Its preadaptive bases—motor control and speech," *J. Hum. Evol.* **14**, 657–668.
- Lubker, J. (1981). "Temporal aspects of speech production: anticipatory labial coarticulation," *Phonetica* **38**, 51–65.
- Lubker, J., and Gay, T. (1982). "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables," *J. Acoust. Soc. Am.* **71**, 437–448.
- Lubker, J., McAllister, R., and Carlson, J. (1975). "Labial coarticulation in Swedish," *Proceedings of the Speech Communications Seminar* (Almqvist and Wiksell, Stockholm).
- Macken, M., and Barton, D. (1980). "The acquisition of the voicing contrast in English: a study of voice-onset time in word-initial stop consonants," *J. Child Lang.* **7**, 41–74.
- Menyuk, P., and Menn, L. (1979). "Early strategies for the perception and production of words and sounds," in *Language Acquisition*, edited by P. Fletcher and M. Garman (Cambridge U. P., Cambridge).
- Miles, F., and Evarts, E. (1979). "Concepts of motor organization," *Ann. Rev. Psychol.* **30**, 327–362.
- Nittrouer, S. (1985). "The role of coarticulation in the perception and production of speech by young children (3 to 7 years)," unpublished, Doctoral dissertation, The City University of New York.
- Ohde, R., and Sharf, D. (1977). "Order effect of acoustic segments of VC and CV syllables on stop and vowel identification," *J. Speech Hear. Res.* **20**, 543–554.
- Ohman, S. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," *J. Acoust. Soc. Am.* **39**, 151–168.
- Polit, A., and Bizzi, E. (1978). "Processes controlling arm movements in monkeys," *Science* **201**, 1235–1237.
- Sharkey, S., and Folkins, J. (1985). "Variability of lip and jaw movements in children and adults: implications for the development of speech motor control," *J. Speech Hear. Res.* **28**, 3–15.
- Soli, S. (1981). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," *J. Acoust. Soc. Am.* **70**, 976–984.
- Turnbaugh, K., Hoffman, P., Daniloff, R., and Absher, R. (1985). "Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers," *J. Acoust. Soc. Am.* **77**, 1256–1257.
- Watkin, K., and Fromm, D. (1984). "Label coordination in children: Preliminary considerations," *J. Acoust. Soc. Am.* **75**, 629–632.
- Winitz, H., Scheib, M., and Reeds, J. (1972). "Identification of stops and vowels for the burst portion of /p t k/ isolated from conversational speech," *J. Acoust. Soc. Am.* **51**, 1309–1317.
- Yeni-Komshian, G., and Soli, S. (1981). "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation," *J. Acoust. Soc. Am.* **70**, 966–975.